BRILL

Multisensory Research 27 (2014) 337–357

MULTISENSORY RESEARCH

brill.com/msr

Visual Objects in the Auditory System in Sensory Substitution: How Much Information Do We Need?

David J. Brown [1,2,*], **Andrew J. R. Simpson** [3] **and Michael J. Proulx** [1,*]

[1] Crossmodal Cognition Lab, Department of Psychology, 2 South, University of Bath, Bath, BA2 7AY, UK

[2] School of Biological and Chemical Sciences, Queen Mary University of London, UK

[3] Centre for Digital Music, Queen Mary University of London, UK

Received 12 October 2013; accepted 1 October 2014

Abstract

Sensory substitution devices such as The vOICe convert visual imagery into auditory soundscapes and can provide a basic 'visual' percept to those with visual impairment. However, it is not known whether technical or perceptual limits dominate the practical efficacy of such systems. By manipulating the resolution of sonified images and asking naïve sighted participants to identify visual objects through a six-alternative forced-choice procedure (6AFC) we demonstrate a 'ceiling effect' at 8×8 pixels, in both visual and tactile conditions, that is well below the theoretical limits of the technology. We discuss our results in the context of auditory neural limits on the representation of 'auditory' objects in a cortical hierarchy and how perceptual training may be used to circumvent these limitations.

Keywords

Sensory substitution, blindness, visual impairment, auditory, object recognition, cross-modal, The vOICe

1. Introduction

Visual impairment affects 285 million people worldwide with 39 million of these legally blind, defined by a visual acuity of less than 20/400 or visual field loss to less than 10° (Pascolini and Mariotti, 2012). While a proportion of cases can be treated through surgical procedures such as the removal of

* To whom correspondence should be addressed. E-mail: d.brown@bath.ac.uk; m.j.proulx@bath.ac.uk

DOI:10.1163/22134808-00002462

cataracts, the development of compensatory techniques is essential for providing a basic visual percept for non-treatable patients. These techniques can be divided into invasive and non-invasive. Invasive techniques involve electrodes implanted in the eye (epi-retinal, sub-retinal and suprachroidal) (Benav *et al.*, 2010; Eickenscheidt *et al.*, 2012; Fujikado *et al.*, 2011; Keseru *et al.*, 2012; Weiland *et al.*, 2005; Zrenner *et al.*, 2011), optic nerve (Chai *et al.*, 2008a, b; Veraart *et al.*, 2003) or cortex (Brindley and Lewin, 1968a, b; Dobelle and Mladejovsky, 1974; Dobelle *et al.*, 1974; Normann *et al.*, 1999; Schmidt *et al.*, 1996).

In the case of retinal implantation, assuming that all implanted electrodes contact the proper retinal cells, state of the art technology incorporating 100 channels provides a theoretical working resolution equivalent to 10 × 10 pixels. However, the simulations of Weiland and colleagues (2005) have suggested that up to 1000 electrodes (e.g., around 30 × 30 pixels) would be necessary for visual processes such as face recognition or text reading. This is supported by Li *et al.*s evaluation of object recognition with retinal implants, which implied an upwards ceiling effect at 24 × 24 pixels (Li *et al.*, 2012).

Non-invasive compensatory techniques rely on technology and neural plasticity to transmit information usually attributed to an impaired sense *via* a neural network of an unimpaired modality. This 'sensory substitution' generally substitutes for impaired vision with the substituting modality being touch (Bach-y-Rita, 2004; Bach-y-Rita and Kercel, 2003; Bach-y-Rita *et al.*, 1969; Danilov and Tyler, 2005; Danilov *et al.*, 2007), or audition (Abboud *et al.*, 2014; Capelle *et al.*, 1998; Meijer, 1992).

The sensory substitution device (SSD) is a three-component system: a sensor (camera) to record information, an algorithm (on PC or smartphone) to convert it, and a transmitter (headphones or tactile array) to relay converted information back to the user. Perceptual resolution, or acuity, of visual-to-tactile (VT) devices are constrained by the distribution of touch receptors at the point of contact (back, fingers, tongue) resulting in low resolutions ranging from simple 10 × 10 systems to the 20 × 20 electrode Brainport (Bach-y-Rita, 2004; Bach-y-Rita *et al.*, 1969; Chebat *et al.*, 2007; Danilov and Tyler, 2005; Sampaio *et al.*, 2001).

Unlike VT devices, visual-to-auditory sensory substitution devices (VA) are not constrained by the density of surface area receptors but instead exploit the wide frequency resolution of the cochlea and the large dynamic range of the auditory nerve. This allows for a much higher theoretical and functional resolution (Haigh *et al.*, 2013; Striem-Amit *et al.*, 2012). As with VT SSDs, resolution varies amongst VA devices. For example, the Prosthesis for Substitution of Vision by Audition (PSVA) has dual resolution function with an 8 × 8 pixel grid of which the four central pixels are each replaced by four smaller ones. The 60 large pixels in the periphery and 64 smaller central pixels (fovea)

give the PSVA a functional resolution of 124 pixels (Capelle *et al.*, 1998). The VA device used in the experiments reported here, The vOICe (Meijer, 1992), which has been used to demonstrate auditory object recognition and localisation (Auvray *et al.*, 2007; Brown *et al.*, 2011; Proulx *et al.*, 2008), utilises a 176×64 pixel array for a functional resolution of up to 11 264 pixels.

This leads to the question: do such systems exhibit ceiling effects in object recognition performance similar to those reported using invasive systems (Li *et al.*, 2012)? The source of such limits on performance can arise at multiple points along the neural pathways processing such information. Many studies of trained users of The vOICe and other SSDs have shown neural activity in brain areas commonly thought of as visual. The sensory modality being stimulated (such as the auditory system) is also stimulated, and likely relays the information to the visual system. Due to the necessary transduction of sensory information in the stimulated modality (such as auditory cortex) before being later processed by the target modality (such as visual cortex), it is fundamental to understand how the capacity of the auditory system impacts the information available for further computations. In auditory–visual substitution, the features of a two-dimensional image which represent an object are encoded as independent spectro-temporal modulations within a complex acoustic waveform (Meijer, 1992). Such acoustic features are encoded independently in the peripheral auditory system and object-based representations emerge in primary auditory cortex (Ding and Simon, 2012; Mesgarani and Chang, 2012; Shamma *et al.*, 2011; Teki *et al.*, 2013). Auditory cortex maintains a two-dimensional topographic map of frequency (Humphries *et al.*, 2010) and modulation-rate (Barton *et al.*, 2012) that are the so-called tonotopic and periodotopic axes, where individual regions on the map independently represent sound features occurring at a specific frequency and modulation rate (Barton *et al.*, 2012; Simon and Ding, 2010; Xiang *et al.*, 2013). It is thought that auditory objects are formed, in cortex, according to temporal coherence between these independently-coded acoustic features (Shamma *et al.*, 2011; Teki *et al.*, 2013).

The representation of spectro-temporal modulation is increasingly rate-limited in the ascending auditory pathway. Phase-locking on the auditory nerve is limited to around 4000 Hz (Joris *et al.*, 2004). By midbrain (inferior colliculus) this limit is reduced to around 300 Hz (Baumann *et al.*, 2011; Joris *et al.*, 2004) and by primary auditory cortex it is further reduced to around 30 Hz (Barton *et al.*, 2012). In superior temporal gyrus (part of Wernicke's speech area), this limit is further reduced to <16 Hz in the object-based representation of speech (Pasley *et al.*, 2012), which limits coincide with those established in human psychoacoustic studies (Simpson and Reiss, 2013; Simpson *et al.*, 2013).

Therefore, different stages of the auditory pathway provide different limits on the visual-sensory substitution problem, where the information encoded in the rendering of the visual image is encoded with increasingly coarse temporal features as it ascends. This is consistent with a reverse-hierarchy theory of multisensory perception and perceptual learning (Proulx *et al.*, 2012), where primary sensory areas provide greater specificity, and higher order areas provide perception at a glance (Ahissar and Hochstein, 2004; Ahissar *et al.*, 2009). If auditory objects are pre-requisite in auditory–visual substitution, this limit is placed earliest at primary auditory cortex. If auditory objects are further refined in higher cortical areas implicated in speech processing, this limit is further strengthened.

These postulations provide testable hypotheses. The image-to-sound rendering system (Meijer, 1992) breaks the visual image into arbitrary pixel sizes which correspond to a resampling of the acoustic modulations by which the image is represented. Shannon–Nyquist sampling theory dictates that the fastest modulations captured are at half the sample (in this case pixel) rate. By varying the pixel resolution of the rendered image it is possible to alter the upper limit (of modulations captured) in a way that is equivalent to the various limits seen on the auditory pathway. If object recognition performance is limited by modulation processing in primary auditory cortex, there should be ceiling effects seen at pixel sampling rates of around 50–60 Hz (giving a cut-off frequency of 25–30 Hz) equivalent to 16×16 pixel visual object (Fig. 1). If performance is limited by higher cortical processing (in speech related areas) then ceiling effects may be seen at even lower pixel (8×8) rates of around 20–30 Hz (giving a cut-off frequency of 10–15 Hz).

The frequency range and temporal length of the sonified stimulus may also be a factor in object recognition. Wright *et al.* (2010) demonstrated generalization to untrained frequencies but not temporal intervals (Wright *et al.*, 1997, 2010) and while evidence shows increased complexity in sonified images increases the breadth of generalization to untrained temporal features (Brown and Proulx, 2013) the extended time course for the latter implies a dominance of frequency components. We therefore categorized our test stimuli into 'short' with a wide frequency range ($M = 3951$ Hz) and short temporal length ($M = 758$ ms) and 'long' with a narrow frequency range ($M = 2280$ Hz) and long temporal length ($M = 951$ ms) — see Fig. 2. This allows us to evaluate whether there is dominance of the spectral (frequency) or temporal (signal length) features of the algorithm in object recognition.

A second stimulus consideration was the use of both visual and tactile objects. The target population for SSD's are those with visual impairment rendering the association between soundscape and visual object meaningless. Our reasoning behind the visual component of the task was that the participants were sighted and naïve to the device. In attempting to demonstrate a proof of

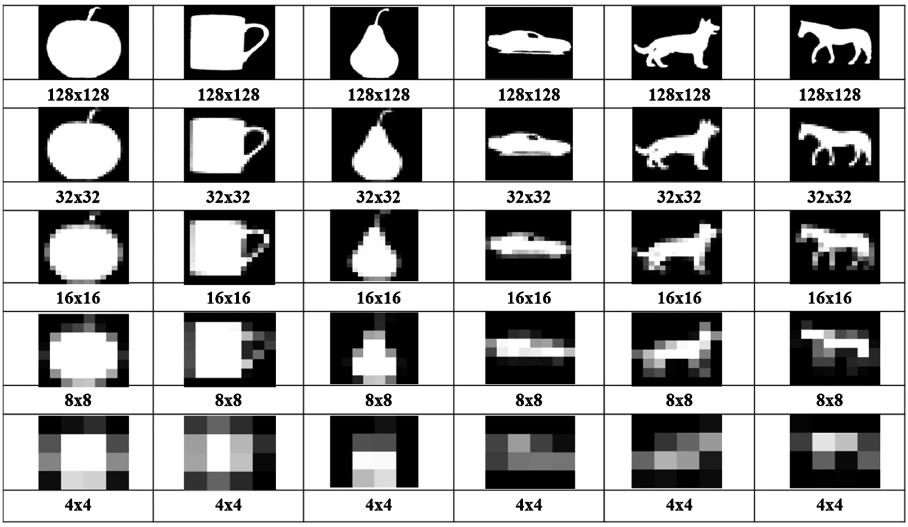

Figure 1. Visual representation of the sonified objects used in the test phases of the experiment. Objects presented to the participant (visually or haptically) were always at the 128 × 128 resolution. The objects at 32 × 32, 16 × 16, 8 × 8, 4 × 4 resolution were sonified using The vOICe and presented as auditory soundscapes only. The participants were never exposed to the visual or tactile objects at the reduced resolutions.

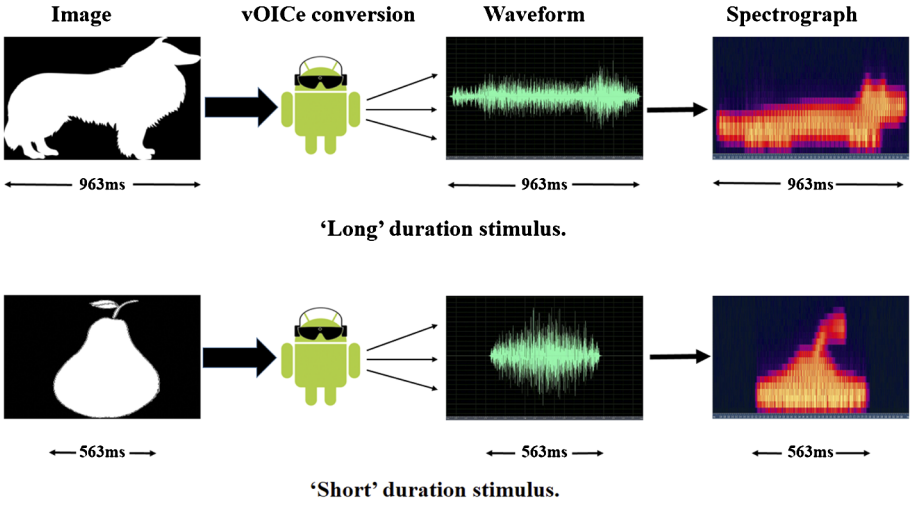

Figure 2. The sonification of one 'long' category object and one 'short' category object. The original visual image is shown along with the waveform and spectrograph of the sonified object. This figure is published in colour in the online version.

concept it seemed logical to train in a familiar modality (vision) for relative simplicity, and a modality relevant to application (tactile).

To evaluate the level of information required for object recognition we used a six alternative forced choice procedure (6AFC) in which listeners had to pair a variously degraded soundscape from The vOICe SSD with one of six 2D objects presented visually or haptically. In the training stage the procedure was similar except all soundscapes were of the full object (i.e., not degraded), it was a 4AFC, and there was post-trial feedback.

Our rationale is threefold. First to evaluate the minimal level of information required for successful object recognition in VA SSD. Based on comparable studies with retinal implants and the visual information displayed in Fig. 1 we predict a ceiling effect at either 8×8 or 16×16 pixels after which an increase in resolution will not elicit superior performance. Secondly to utilise a behavioural paradigm to assess where in the auditory hierarchy resolution-based objects are processed. For the larger of our predicted ceiling effects we hypothesise auditory object recognition in primary auditory cortex, with lower ceiling effects further up the auditory pathway. Finally we were interested in whether recognition would be better for stimuli with a 'short' duration and wide frequency range than for those with a 'long' duration and narrow frequency range. This is exploratory and so we make no prediction in either direction for this assessment.

2. Method

2.1. Participants

We recruited 19 undergraduate students (12 female) from 18 to 28 years of age ($M = 20.42$, SD $= 3.22$) from Queen Mary University of London. Two participants withdrew from the study after the training session, due to personal reasons, so 17 participants (ten female) age range 18 to 28 ($M = 20.71$, SD $= 3.29$) took part in the test phase. All participants reported normal or corrected vision and normal hearing. 16 (training) and 14 (test) were right-handed. The study was approved by Queen Mary university of London ethics Committee REC/2009 and all participants provided written consent prior to the study onset. Remuneration was *via* the undergraduate course credit scheme with an additional £0.05 per correct response in the test phases.

2.2. Materials

'Auditory' stimuli were created using The vOICe (Meijer, 1992), Adobe Audition 3 and Adobe Photoshop CS3 (see stimulus design below). The script was run in E-Prime 2.0 (Psychology Software Tools, Pittsburgh, PA, USA) on a Windows 7 desktop PC. All auditory signals were transmitted *via* Sennheiser HD555 full ear headphones. Images to be sonified were obtained from EST 80

image set (Max Planck Institute, Germany) and Clipart. The blindfold was the Mindfold (Mindfold Inc. Tucsan, AZ, USA).

2.3. Stimulus Design

Images were transformed to soundscapes using The vOICe's image sonification feature at default settings (1 s scan rate, normal contrast, foveal view off). Visual images were white on a black background with a 1 s duration on the x-axis and a 500–5000 Hz frequency range on the y-axis. Tactile stimuli were created by cutting the object shape (white area) from 5 mm foam board and attaching this to 90 × 55 mm card backgrounds, For the training days there were 40 different objects in total (34 on day one).

2.3.1. Test Day Stimuli — Object Resolution and Categorization

During the test phases only six visual and six tactile stimuli were presented to the participant. These were all at 128 × 128 pixels. These visual images were manipulated in Adobe Photoshop to produce variants at four pixel resolutions (32 × 32, 16 × 16, 8 × 8, 4 × 4) and then sonified (Fig. 1). Hence the tactile or visual objects were always at 128 × 128 pixel resolution while the soundscapes were at various lower resolutions subdivided into two categories based on the temporal and spectral features of the rendered soundscape. Three objects were 'long' on the x-axis but narrow on the y-axis (car, dog, horse) with the other three relatively 'short' on the x-axis but with broad range of frequencies on the y-axis (apple, pear cup). When sonified this resulted in either long, spectrally sparse or short spectrally dense signals, as shown in Fig. 2.

2.4. Procedure

2.4.1. Training Day One

Participants were shown a PowerPoint presentation explaining The vOICe algorithm, including worked audio-visual examples and an explanation of the experimental task. For each task trial participants listened to a soundscape (repeated four times) while looking at a blank screen. The soundscapes were then repeated accompanied by four numbered images on the screen. The participant indicated, using 1–4 on a numeric keypad, which image had been sonified to create the soundscape. The soundscape could be repeated by pressing 'R' and visual feedback was given post response in a correct/incorrect format prior to onset of the next trial.

There were 32 trials in each of two blocks. Each block had four categories of trial, varying in difficulty based on object features. For example, in the first eight trials the correct object varied greatly from the three alternates. For the second set there were two obviously different alternates and so on. The trials alternated between filled and empty objects (object outline only) to evaluate The vOICe's edge enhancement feature in early stage training. The second day one training phase replicated the first, except that images were sonified at a 2 s

scan rate. For the final two blocks on training day one the participants were blindfolded and undertook a similar 4AFC procedure involving associations to be made between the soundscapes and the haptically explored tactile objects. Responses and requests for repeat presentations were instigated by the experimenter. Tactile blocks were completed after the visual ones for all participants. Otherwise all presentation orders were counterbalanced.

2.4.2. Training Day Two

The second training day was a replication of day one (minus Powerpoint presentation) utilising different 4AFC's, and reversing the procedure so the participant was presented with one object (visual or tactile) and 4 soundscapes (each repeated 4 times). The six test day objects (at 128 × 128 pixels) were introduced into this session, although the participants were unaware these were the test day objects. 1 or 2 s scan rate order was counterbalanced across days. After the second training day, participants who had a $\geqslant$50% correct response rate (based on a pilot study with different participants) were invited to return for the test phases.

2.4.3. Test Day One

Methodologically this was similar to the training phases with a number of alterations. Firstly there were six presented objects in each trial (6AFC) with the same six objects being presented for each trial. Secondly there was no post-trial feedback. Thirdly, there were 72 trials in each block of the visual test phase and 36 in each tactile block. Participants were given six visual or haptic objects and required to match the soundscape to one of them, either by responding 1–6 on the keyboard (visual) or verbalising a response (tactile). Again a repeat feature was available to listen to the soundscape again prior to responding.

2.4.4. Test Day Two

As with the training days this was a reversal in procedure. For each trial participants were presented with six soundscapes (each repeated four times) and one visual or tactile object. The task was to indicate which of the six objects had been sonified. As in test day one, there was no post trial feedback. The order of test days was counterbalanced across participants but the visual-soundscape association was always performed first.

3. Results

The primary objective of the experiment was to evaluate auditory object recognition, at increasingly coarse resolutions, using a VA SSD. We were also interested in whether the temporal and spectral composition of the stimuli were a factor in successful object recognition, and finally, in the initial training ses-

sions, if empty or filled objects and different duration scan rates would elicit superior performance.

3.1. Object Resolution — Visual/Soundscape

Figure 3 and Table 1 shows performance accuracy (%) as a function of resolution for the visual/soundscape matching condition. The means and standard deviations for each resolution category are displayed in Table 1. While successful recognition was better than the 6AFC chance level of 16.67% for all resolutions ($p < 0.05$), implying successful use of the device irrespective of object resolution, there was a significant difference between the perfor-

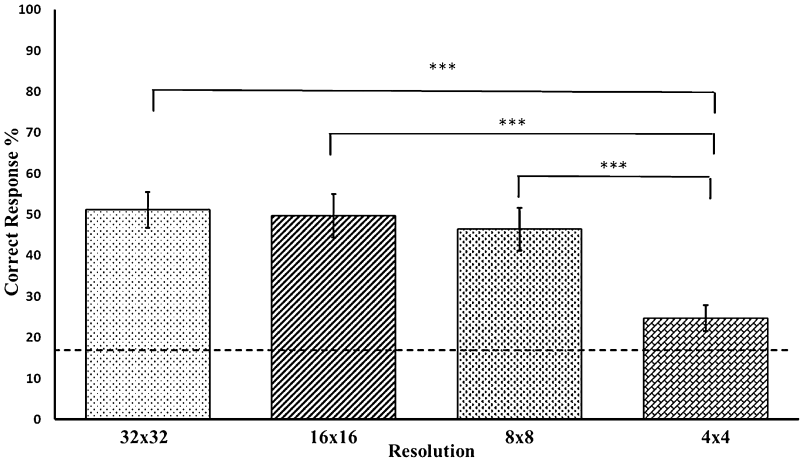

Figure 3. Successful object recognition in the visual-to-auditory → visual matching condition based on object resolution. The dashed line represents what would be expected by chance. Contrast bars indicate significant differences between conditions with error bars displaying SEM. * for $p < 0.05$, ** for $p < 0.01$, and *** for $p < 0.001$.

Table 1.
Mean correct scores (%) and standard deviations for the correct responses in the visual-to-auditory → visual matching and the visual-to-auditory → tactile matching tasks. Percentages are given for the different resolutions and totals for each modality

Resolution	Visual		Tactile	
	Mean %	SD	Mean %	SD
32 × 32	51.14	18.08	56.62	23.54
16 × 16	49.67	21.89	48.16	20.46
8 × 8	46.41	21.40	35.64	16.31
4 × 4	24.67	13.17	20.39	12.13
Total	42.61	17.15	39.71	14.67

mance in the four categories ($F[3, 48] = 28.686$, $p < 0.001$, $\eta p^2 = 0.642$). Bonferroni-corrected planned contrasts showed that the highest resolution, 32 × 32 ($M = 51.14\%$, SD = 18.08) was better recognised compared to 4 × 4 ($M = 24.67\%$, SD = 13.17) with a mean difference (Md) of 26.47% (95% CI [17.69, 35.25], $p < 0.001$), but not compared to 16 × 16 ($M = 49.67\%$, SD = 21.89), ($p = 0.988$) or 8 × 8 ($M = 46.41\%$, SD = 21.40), ($p = 0.556$). Performance on the 16 × 16 resolution was superior to 4 × 4 ($M = 25.00\%$, 95% CI [12.99, 37.01], $p < 0.001$) but not 8 × 8 ($p = 1.00$). The final contrast demonstrated that recognition of stimuli at 8 × 8 was significantly better than 4 × 4 ($M = 21.73\%$, 95% CI [11.59, 31.87], $p < 0.001$) and 8 × 8.

3.2. Object Resolution — Tactile/Soundscape

Figure 4 and Table 1 show the results for the tactile/soundscape matching condition. Performance was above chance for the three higher resolutions but, unlike the visual matching condition, not for the 4 × 4 ($t[16] = 1.269$, $p = 0.223$, $d = 0.635$). There was a significant main effect of resolution on tactile/soundscape matching ($F[3, 48] = 23.019$, $p < 0.001$, $\eta p^2 = 0.590$) with the mean differences in percentage for 4 × 4 ($M = 20.39\%$, SD = 12.13) soundscapes poorly matched compared to 32 × 32 ($M = 56.62\%$, SD = 23.54), ($M = 36.23\%$, 95% CI [21.09, 51.37], $p < 0.001$), 16 × 16 ($M = 48.16\%$, SD = 20.46), ($M = 27.77\%$, 95% CI [12.94, 42.60], $p < 0.001$), and 8 × 8 ($M = 35.64\%$, SD = 16.31), ($M = 15.25\%$, 95% CI [4.87, 25.63], $p = 0.003$) demonstrating that recognition of the lowest reso-

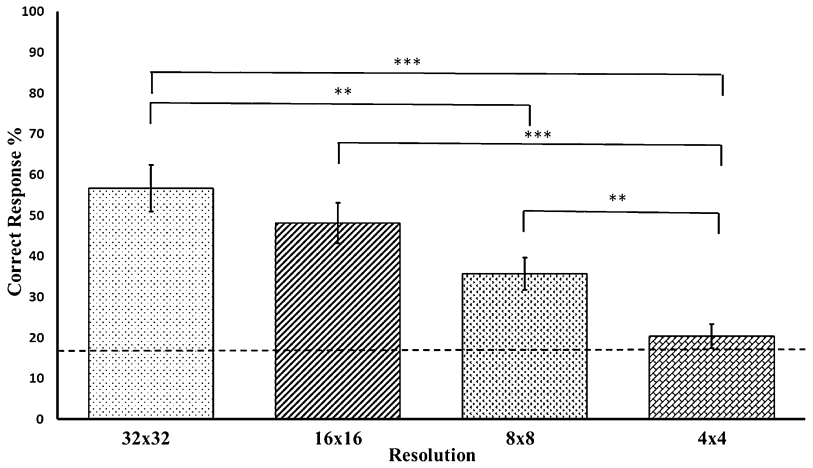

Figure 4. Successful object recognition in the visual-to-auditory → tactile matching condition based on object resolution. The dashed line represents what would be expected by chance. Contrast bars indicate significant differences between conditions with error bars displaying SEM. * for $p < 0.05$, ** for $p < 0.01$, and *** for $p < 0.001$.

lution soundscapes were difficult irrespective of object modality. Unlike the visual matching condition where performance varied little above the ceiling effect of the 8 × 8 trials, there was a distinct advantage for the higher resolution objects in the haptic condition: recognition in 32 × 32 was better than 8 × 8 ($M = 20.98\%$, 95% CI [4.61, 37.34], $p = 0.008$), and 16 × 16 although not quite at significance for the latter ($p = 0.059$).

T-tests were performed to compare 'visual' and tactile conditions for each resolution. Tactile performance at the highest resolution was better than its visual counterpart although non-significant ($p = 0.113$). Visual recognition was superior for the other three resolutions, significant at 8 × 8 ($t[16] = 3.272$, $p = 0.005$, $d = 0.794$) but not for 16 × 16 ($p = 0.740$) or 4 × 4 ($p = 0.118$).

3.3. Object Type

The secondary analysis considered object recognition as a function of stimulus type. Three objects were classified as 'long' and the other three as 'short' based on the temporal duration of the signal. The latter group also were composed of a wider range of frequencies compared to the former. Figure 5 and Table 2 show the results for the individual objects. Collapsed across the two categories (long + short) there was no significant difference between 'long' ($M = 44.20\%$, SD = 17.34) and 'short' ($M = 41.42\%$, SD = 19.13) in the visual matching task ($t[16] = 0.969$, $p = 0.347$, $d = 0.235$). In the haptic condition, recognition for objects in the 'short' category ($M = 44.51\%$, SD =

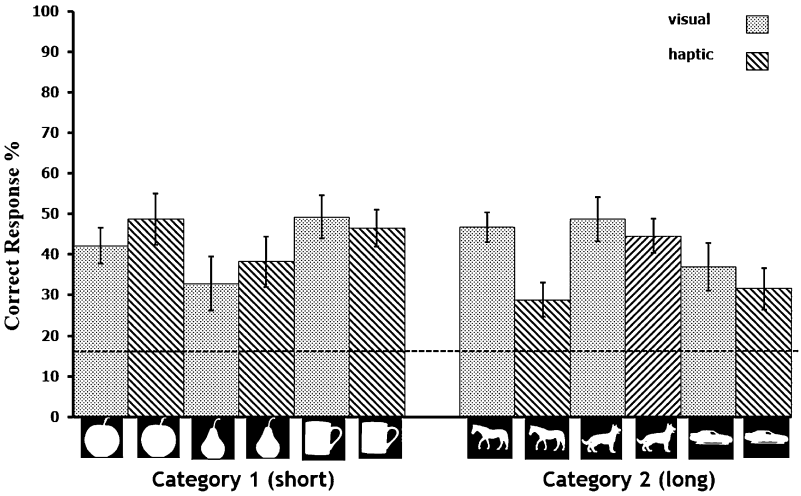

Figure 5. Successful object recognition for each individual object in both visual-to-auditory → visual matching and visual-to-auditory → tactile matching. Objects are categorised into 'long' and 'short' conditions based on the temporal length of the active part of the soundscape. The dashed line indicates what would be expected by chance with error bars displaying SEM.

Table 2.
Mean correct scores (%) and standard deviations for individual object recognition. Percentages are given for each object and a total for both the 'long' and 'short' conditions

Object	Visual matching		Tactile matching	
	Mean %	SD	Mean %	SD
Pear	32.84	27.20	38.24	25.55
Apple	42.16	18.39	48.82	25.95
Cup	49.27	21.71	46.47	18.69
'Short' category	41.42	19.13	44.51	17.91
Dog	48.78	22.62	44.61	17.66
Horse	46.81	15.20	28.88	17.37
Car	37.01	24.20	31.55	20.86
'Long' category	44.20	17.34	35.29	13.15

17.91) was superior to those in the 'long' category ($M = 35.29\%$, SD = 13.15); ($t[16] = 3.417$, $p = 0.004$, $d = 0.860$).

To find the source of these differences, the individual objects were analysed looking at both intra- and intergroup comparisons. In the visual condition there was an overall main effect of object type ($F[5, 80] = 3.543$, $p = 0.006$, $\eta p^2 = 0.181$) with intragroup differences between cup *versus* pear (short) ($p = 0.014$) and dog *versus* car (long) ($p = 0.009$). Intergroup contrasts demonstrated performance differences for dog *versus* pear ($p = 0.006$), horse *versus* pear ($p = 0.034$) and borderline effect for cup *versus* car ($p = 0.057$).

There was also a main effect of stimulus type in the tactile/soundscape matching condition ($F[5, 80] = 4.053$, $p = 0.002$, $\eta p^2 = 0.202$) with contrasts showing intracategory differences for dog *versus* horse ($p = 0.026$), dog *versus* car ($p = 0.02$) and a borderline result in the 'short' apple *versus* pear ($p = 0.067$). Intercategory contrasts in this condition were significant for cup *versus* horse ($p = 0.007$), cup *versus* car ($p = 0.034$), apple *versus* car ($p = 0.003$), apple *versus* horse ($p = 0.003$) and borderline pear *versus* horse ($p = 0.055$).

3.4. Procedure Comparison

The final analysis in the test phase contrasted performance over the two test sessions. Training effects would suggest superior performance for day two. Conversely we found overall performance on the second day ($M = 39.59\%$, SD = 18.15) to be worse than day one ($M = 42.72\%$, SD = 14.39) although not reaching significance ($t[16] = 1.447$, $p = 0.167$, $d = 0.351$). If this comparison is made with the data divided by stimulus type, visual performance on day one ($M = 45.18\%$, SD = 16.66) is significantly better than for day two

($M = 40.03\%$, SD $= 18.63$) ($t[16] = 2.492$, $p = 0.024$, $d = 0.604$) but this is not found for the tactile condition ($t[16] = 0.333$, $p = 0.744$, $d = 0.081$). Our two test days differed in the presentation of the 6AFC. On day one the participant was presented with six visual/haptic objects and one soundscape. This method of presentation is clearly less problematic to the listener than if given one object and six soundscapes, as on day 2.

3.5. Training

The structure and stimuli in the training regime allowed us to evaluate device settings in naïve users. Objects were either filled, where the whole object was white, or empty, where only the object outline was in white. Device scan rates were either 1 s or 2 s to give four stimulus conditions. Table 3 displays the mean performance for these conditions. For visual/soundscape matching analysis of variance showed a main effect of performance as a function of condition ($F[3, 54] = 4.366$, $p = 0.008$, $\eta p^2 = 0.195$). Bonferroni-corrected contrasts found no significant pairwise comparisons. However trends suggested that the 1 s filled stimuli were poorly recognised compared to 2 s filled ($p = 0.059$), and 2 s empty ($p = 0.061$) implying that the time scan may have had some effect. Analysis on this data collapsed into 'time scan' and 'filled/empty' groups showed that performance on the 2 s scan rate ($M = 64.31\%$, SD $= 14.22$) was superior to its 1 s counterpart ($M = 57.81\%$, SD $= 10.90$), ($t[18] = 2.914$, $p = 0.009$, $d = 0.668$) but not reaching significance for filled ($M = 62.66\%$, SD $= 12.53$) *versus* empty ($M = 59.46\%$, SD $= 12.81$) shapes ($t[18] = 1.438$, $p = 0.168$, $d = 0.330$).

These results contrast with those of Brown *et al.* (2011) who evaluated different vOICe device settings in object recognition and found no significant

Table 3.
Mean correct scores (%) and standard deviations for the different conditions in the training phases of the experiment

	Visual matching		Tactile matching	
	Mean %	SD	Mean %	SD
1 s filled	60.86	11.48	67.43	11.66
1 s empty	54.77	14.60		
2 s filled	64.47	15.84		
2 s empty	64.14	14.71		
Filled total	62.67	12.53		
Empty total	59.46	12.81		
1 s total	57.81	10.90		
2 s total	64.31	14.22		

advantage for the 2 s scan speed over the 1 s. This can be attributed to paradigm differences with the former using the device in real time with real objects at multiple perspectives and the later utilising sonified two-dimensional images. There is clearly an advantage to a slower scan speed if the objects are simple and the soundscape consistent over time.

4. Discussion

In this study we evaluated object recognition performance in naïve users of a VA SSD, The vOICe. Images, and their soundscapes, were manipulated by pixel resolution to ascertain the minimal amount of visual/tactile/soundscape information that is needed for successful recognition. As secondary considerations we looked at the spectral/temporal composition of the stimuli and presentation order within the 4AFC as factors in recognition, and replicated various device settings in training to assess for any preference. The results demonstrate a lower ceiling effect of 8 × 8 (64) pixels in both the visual-VA and tactile-VA conditions for object resolution. While this is informative for structuring effective training regimes it also allows postulations on cortical representation of sonified objects.

In both invasive and non-invasive SSD systems the central 'visual' system (i.e., cortex) is implicated in processing of visual objects. Imaging studies have demonstrated the recruitment of 'visual' areas in VA SSD use, even in naïve users (Arno *et al.*, 2001; Poirier *et al.*, 2006) with transcranial magnetic stimulation (TMS) to visual cortex impeding pattern recognition tasks using SSD's (Collignon *et al.*, 2007). Output from The vOICe also shows activation in areas of lateral occipital cortex, an area not associated with auditory input, implying that the 'auditory' signal from the device is not only processed in the auditory pathway (Amedi *et al.*, 2007; Plaza *et al.*, 2012). This is further corroborated by evidence of a correlation between musical ability and performance using a VA SSD (Haigh *et al.*, 2013) This leads to the further question: are the limits of such systems to be found in auditory or visual neural circuits?

If auditory object recognition is a limiting factor, then information processing in primary auditory cortex is crucial; phase locking in auditory cortex is limited to around 30 Hz, thus we would expect a ceiling effect at the 16 × 16 image resolution (Barton *et al.*, 2012). However, the ceiling effect at 8 × 8 pixels suggests that object recognition is instead processed further up the auditory pathway, such as in the superior temporal gyrus (STG) where phase locking is reduced to $<$16 Hz. This is consistent with performance by higher cortical representations optimized for speech processing (Pasley *et al.*, 2012). The implications of this are that the pre-lexical higher-cortical object-based representation constitutes the ultimate token that allows the listeners to recognize a rendered object and places strict limits on the potential success of

the substitution system, and subsequent processing in visual or supramodal cortical areas. This does not mean that these limits, as implicit in the use of a higher cortical speech processor, negate the viability of SSD's and indeed may be circumvented by building cross-modal networks at the earlier level of primary cortex (or even midbrain). Extensive training and learning on the devices might, *via* synaptic plasticity, produce cross-modal networks capable of exploiting earlier, wider-bandwidth representations thus bypassing the limitations of the speech processor. Indeed recruitment of higher multisensory processing cortical areas, such as the STG, may be key in allowing information transfer between primary sensory areas thus giving rise to higher fidelity information processing and even visual imagery in some long term device users (Proulx *et al.*, 2014; Ward and Meijer, 2010).

The ceiling effect at 8×8 draws interesting comparisons with Weiland and colleagues (2005) simulations for retinal implants. Their estimation of a 30×30 electrode/pixel array being a requisite for face recognition and text reading may be overstated. While noting we are comparing invasive and non-invasive techniques and different paradigms, the 8×8 ceiling with minimal improvement at higher resolutions, implies the brain can extract enough salient information from coarse SSD input for effective object/pattern recognition.

As well as being affected by resolution, object recognition was also influenced by stimulus type (visual/tactile), stimulus features (long/short temporal length), and task procedure. The soundscapes in both the visual and haptic matching tasks were identical and therefore any performance differences can be attributed to modality specific difficulties in object identification rather than processing of the SSD signal. Unsurprisingly, visual/soundscape matching was more successful than the haptic counterpart. All participants were sighted and therefore their primary modality for 'everyday' object recognition is vision.

Visual object recognition utilises a number of cues such as shape, luminance, depth, motion, shading and colour which are processed in parallel to allow a rapid identification of the object, in usually about 1 s (Martinovic *et al.*, 2008). Object recognition *via* haptics is less rapid and usually serial (Overvliet *et al.*, 2007) as individual object features have to be explored sequentially, committed to memory, and mentally reassembled to give a percept of the object (Craddock and Lawson, 2008). If time-based haptic exploration is slower, and logic dictates that larger objects require more exploration time, and then the advantage for 'short' objects in the haptic condition, compared to 'long', is understandable. This would be salient if a time limit was placed on the trial forcing object identification to be rapid. In the present experiment there was no 'official' time limit placed on the task but having completed the more rapid 'visual' task first participants may have responded in the haptic task at a speed familiar to the procedure.

The procedure was certainly a main effector on the results. On test day one all stimuli in the trial (all visual/haptic objects plus one repeated soundscape) are presented to the participant 'online' simultaneously for the duration of the trial. Visual–auditory feature matching, and saliently, comparison between features of different objects can be done quickly with little memory load. On test day two the visual/haptic object is available for the trial duration but the six soundscapes are sequentially presented. Feature matching, particularly comparisons, requires memory load in the retention and recall of previous soundscapes. While all six tactile objects on day one are 'available' to the participant for the duration of the trial, haptic exploration is still serial as all objects cannot be haptically explored concurrently.

The level and duration of visual impairment in the target group may also be influential on the ability to use different levels of resolution in sensory substitution. While the data collected on sighted participants may be extrapolated to inform sensory augmentation (e.g., expansion of the FOV), where the device is not substituting for an impaired sense but providing additional information to a fully functioning perceptual system, processing differences in late, and particularly, congenitally blind participants may elicit different results. Behavioural and neural differences between sighted, late and congenitally blind have been demonstrated for, amongst other things, false memories, the mental number line, and spatial representations (Pasqualotto *et al.*, 2013a, b). Pasquallotto and colleagues found in a spatial task that while sighted and late blind participants showed a preferential use of an object-based or 'allocentric' reference frame, the congenitally blind participants preferred a self-based 'egocentric' reference frame (Pasqualotto *et al.*, 2013b). This corresponds with ideas that at least some visual experience is a requisite of developing multisensory neurons, spatial updating tasks, multisensory integration and higher cognition (Pasqualotto and Proulx, 2012; Reuschel *et al.*, 2012; Wallace *et al.*, 2004). With two algorithm principles coding spatial factors and multisensory integration integral in SSD use, task-based comparisons between the three should feature heavily in future research.

The results of the present study feed directly into theories regarding standardization of working resolutions across devices. SSD's are limited in the information they can convey by their conversion algorithms; that is, three principles can only transmit three aspects of visual perception. One way to overcome this is to utilise numerous SSD's (VT + VA) or a combination of invasive and non-invasive devices. Should we establish a consistent working resolution across devices to develop effective training protocols that maximise the effectiveness of multiple device use? A functional limit (24×24) for basic object recognition has been ascertained for retinal implants (Li *et al.*, 2012). If it holds that successful object recognition can be achieved at lower resolutions in SSD's then this informs on the use of each device in an

invasive/non-invasive combination, i.e., the SSD for fine-grained recognition and the implant for more coarse spatial/navigation information. A final consideration in applying these results to developing training protocols is 'how high a resolution is sufficient/desirable for successful object recognition in sensory substitution?' As stated by Paul Bach-y-Rita:

> "*A poor resolution sensory substitution system can provide the information necessary for the perception of complex images. The inadequacies of the skin (e.g. poor two-point resolution) do not appear as serious barriers to eventual high performance, because the brain extracts information from the patterns of stimulation. It is possible to recognise a face or to accomplish hand–eye coordinated tasks with only a few hundred points of stimulation.*" (Bach-y-Rita and Kercel, 2003, p. 543)

In conclusion we have demonstrated an apparent resolution ceiling effect (8×8 pixels) in which successful object recognition is possible in naïve users of a VA SSD and postulated that in such users the ascending auditory hierarchy may place limitations on such a task. Further research should be undertaken to evaluate how this can be extrapolated to extensively trained users, late and congenitally blind users and situations in 'real time'. A more comprehensive understanding of this would allow us to develop more effective training protocols for sensory substitution and give us a better understanding of the associated brain processes.

Acknowledgements

This work was supported in part by a grant from the EPSRC to MJP (EP/J017205/1) and the EPSRC Doctoral Training Account studentship at Queen Mary University of London to AJRS.

References

Abboud, S., Hanassy, S., Levy-Tzedek, S., Maidenbaum, S. and Amedi, A. (2014). EyeMusic: introducing a "visual" colorful experience for the blind using auditory sensory substitution, *Rest. Neurol. Neurosci.* **32**, 247–257.

Ahissar, M. and Hochstein, S. (2004). The reverse hierarchy theory of visual perceptual learning, *Trends Cogn. Sci.* **8**, 457–464.

Ahissar, M., Nahum, M., Nelken, I. and Hochstein, S. (2009). Reverse hierarchies and sensory learning, *Phil. Trans. R. Soc. B* **364**, 285–299.

Amedi, A., Stern, W. M., Camprodon, J. A., Bermpohl, F., Merabet, L., Rotman, S., Hemond, C., Meijer, P. and Pascual-Leone, A. (2007). Shape conveyed by visual-to-auditory sensory substitution activates the lateral occipital complex, *Nat. Neurosci.* **10**, 687–689.

Arno, P., De Volder, A. G., Vanlierde, A., Wanet-Defalque, M. C., Streel, E., Robert, A., Sanabria-Bohórquez, S. and Veraart, C. (2001). Occipital activation by pattern recognition in the early blind using auditory substitution for vision, *Neuroimage* **13**, 632–645.

Auvray, M., Hanneton, S. and O'Regan, J. K. (2007). Learning to perceive with a visuo-auditory substitution system: localisation and object recognition with 'The vOICe', *Perception* **36**, 416–430.

Bach-y-Rita, P. (2004). Tactile sensory substitution studies, *Coevol. Hum. Potential Converging Technol.* **1013**, 83–91.

Bach-y-Rita, P. and Kercel, S. W. (2003). Sensory substitution and the human–machine interface, *Trends Cogn. Sci.* **7**, 541–546.

Bach-y-Rita, P., Collins, C. C., Saunders, F. A., White, B. and Scadden, L. (1969). Vision substitution by tactile image projection, *Nature* **221**(5184), 963–964.

Barton, B., Venezia, J. H., Saberi, K., Hickok, G. and Brewer, A. A. (2012). Orthogonal acoustic dimensions define auditory field maps in human cortex, *Proc. Natl Acad. Sci. USA* **109**, 20738–20743.

Baumann, S., Griffiths, T. D., Sun, L., Petkov, C. I., Thiele, A. and Rees, A. (2011). Orthogonal representation of sound dimensions in the primate midbrain, *Nat. Neurosci.* **14**, 423–425.

Benav, H., Bartz-Schmidt, K. U., Besch, D., Bruckmann, A., Gekeler, F., Greppmaier, U., Harscher, A., Kibbel, S., Kusnyerik, A., Peters, T., Sachs, H., Stett, A., Stingl, K., Wilhelm, B., Wilke, R., Wrobel, W. and Zrenner, E. (2010). Restoration of useful vision up to letter recognition capabilities using subretinal microphotodiodes, *Conf. Proc. IEEE Eng. Med. Biol. Soc.* **2010**, 5919–5922.

Brindley, G. S. and Lewin, W. S. (1968a). The sensations produced by electrical stimulation of the visual cortex, *J. Physiol.* **196**, 479–493.

Brindley, G. S. and Lewin, W. S. (1968b). The visual sensations produced by electrical stimulation of the medial occipital cortex, *J. Physiol.* **194**, 54–55P.

Brown, D. J., Macpherson, T. and Ward, J. (2011). Seeing with sound? Exploring different characteristics of a visual-to-auditory sensory substitution device, *Perception* **40**, 1120–1135.

Brown, D. J. and Proulx, M. J. (2013). Increased signal complexity improves the breadth of generalization in auditory perceptual learning, *Neural Plast.* 879047. DOI:10.1155/2013/879047.

Capelle, C., Trullemans, C., Arno, P. and Veraart, C. (1998). A real-time experimental prototype for enhancement of vision rehabilitation using auditory substitution, *IEEE Trans. Biomed. Eng.* **45**, 1279–1293.

Chai, X., Li, L., Wu, K., Zhou, C., Cao, P. and Ren, Q. (2008a). C-sight visual prostheses for the blind, *IEEE Eng. Med. Biol. Mag.* **27**(5), 20–28.

Chai, X., Zhang, L., Li, W., Shao, F., Yang, K. and Ren, Q. (2008b). Study of tactile perception based on phosphene positioning using simulated prosthetic vision, *Artif Organs* **32**, 110–115.

Chebat, D. R., Rainville, C., Kupers, R. and Ptito, M. (2007). Tactile-'visual' acuity of the tongue in early blind individuals, *Neuroreport* **18**, 1901–1904.

Collignon, O., Lassonde, M., Lepore, F., Bastien, D. and Veraart, C. (2007). Functional cerebral reorganization for auditory spatial processing and auditory substitution of vision in early blind subjects, *Cereb. Cortex* **17**, 457–465.

Craddock, M. and Lawson, R. (2008). Repetition priming and the haptic recognition of familiar and unfamiliar objects, *Percept. Psychophys.* **70**, 1350–1365.

Danilov, Y. P. and Tyler, M. (2005). Brainport: an alternative input to the brain, *J. Integr. Neurosci.* **4**, 537–550.

Danilov, Y. P., Tyler, M. E., Skinner, K. L., Hogle, R. A. and Bach-y-Rita, P. (2007). Efficacy of electrotactile vestibular substitution in patients with peripheral and central vestibular loss, *J. Vestib. Res.* **17**, 119–130.

Ding, N. and Simon, J. Z. (2012). Emergence of neural encoding of auditory objects while listening to competing speakers, *Proc. Natl Acad. Sci. USA* **109**, 11854–11859.

Dobelle, W. H. and Mladejovsky, M. G. (1974). Phosphenes produced by electrical stimulation of human occipital cortex, and their application to the development of a prosthesis for the blind, *J. Physiol.* **243**, 553–576.

Dobelle, W. H., Mladejovsky, M. G. and Girvin, J. P. (1974). Artifical vision for the blind: electrical stimulation of visual cortex offers hope for a functional prosthesis, *Science* **183**(4123), 440–444.

Eickenscheidt, M., Jenkner, M., Thewes, R., Fromherz, P. and Zeck, G. (2012). Electrical stimulation of retinal neurons in epiretinal and subretinal configuration using a multicapacitor array, *J. Neurophysiol.* **107**, 2742–2755.

Fujikado, T., Kamei, M., Sakaguchi, H., Kanda, H., Morimoto, T., Ikuno, Y., Nishida, K., Kishima, H., Maruo, T., Konoma, K., Ozawa, M. and Nishida, K. (2011). Testing of semichronically implanted retinal prosthesis by suprachoroidal-transretinal stimulation in patients with retinitis pigmentosa, *Invest. Ophthalmol. Vis. Sci.* **52**, 4726–4733.

Haigh, A., Brown, D. J., Meijer, P. and Proulx, M. J. (2013). How well do you see what you hear? The acuity of visual-to-auditory sensory substitution, *Front. Psychol.* **4**, 330.

Humphries, C., Liebenthal, E. and Binder, J. R. (2010). Tonotopic organization of human auditory cortex, *Neuroimage* **50**, 1202–1211.

Joris, P. X., Schreiner, C. E. and Rees, A. (2004). Neural processing of amplitude-modulated sounds, *Physiol. Rev.* **84**, 541–577.

Keseru, M., Feucht, M., Bornfeld, N., Laube, T., Walter, P., Rossler, G., Velikay-Parel, M., Hornig, R. and Richard, G. (2012). Acute electrical stimulation of the human retina with an epiretinal electrode array, *Acta Ophthalmol.* **90**, e1–e8. DOI:10.1111/j.1755-3768.2011.02288.x.

Li, S., Hu, J., Chai, X. Y. and Peng, Y. H. (2012). Image recognition with a limited number of pixels for visual prostheses design, *Artif. Organs* **36**, 266–274.

Martinovic, J., Gruber, T., Hantsch, A. and Muller, M. M. (2008). Induced gamma-band activity is related to the time point of object identification, *Brain Res.* **1198**, 93–106.

Meijer, P. B. L. (1992). An experimental system for auditory image representations, *IEEE Trans. Biomed. Eng.* **39**, 112121.

Mesgarani, N. and Chang, E. F. (2012). Selective cortical representation of attended speaker in multi-talker speech perception, *Nature* **485**(7397), 233–236.

Normann, R. A., Maynard, E. M., Rousche, P. J. and Warren, D. J. (1999). A neural interface for a cortical vision prosthesis, *Vis. Res.* **39**, 2577–2587.

Overvliet, K. E., Smeets, J. B. and Brenner, E. (2007). Parallel and serial search in haptics, *Percept. Psychophys.* **69**, 1059–1069.

Pascolini, D. and Mariotti, S. P. (2012). Global estimates of visual impairment: 2010, *Br. J. Ophthalmol.* **96**, 614–618.

Pasley, B. N., David, S. V., Mesgarani, N., Flinker, A., Shamma, S. A., Crone, N. E., Knight, R. T. and Chang, E. F. (2012). Reconstructing speech from human auditory cortex, *PLoS Biol.* **10**(1), e1001251. DOI:10.1371/journal.pbio.1001251.

Pasqualotto, A. and Proulx, M. J. (2012). The role of visual experience for the neural basis of spatial cognition, *Neurosci. Biobehav. Rev.* **36**, 1179–1187.

Pasqualotto, A., Lam, J. S. and Proulx, M. J. (2013a). Congenital blindness improves semantic and episodic memory, *Behav. Brain Res.* **244**, 162–165.

Pasqualotto, A., Spiller, M. J., Jansari, A. S. and Proulx, M. J. (2013b). Visual experience facilitates allocentric spatial representation, *Behav. Brain Res.* **236**, 175–179.

Pasqualotto, A., Taya, S. and Proulx, M. J. (2014). Sensory deprivation: visual experience alters the mental number line, *Behav. Brain Res.* **261**, 110–113.

Plaza, P., Cuevas, I., Grandin, C., De Volder, A. G. and Renier, L. (2012). Looking into task-specific activation using a prosthesis substituting vision with audition, *ISRN Rehabil.* **2012**, 15. DOI:10.5402/2012/490950.

Poirier, C. C., De Volder, A. G., Tranduy, D. and Scheiber, C. (2006). Neural changes in the ventral and dorsal visual streams during pattern recognition learning, *Neurobiol. Learn. Mem.* **85**, 36–43.

Proulx, M. J., Stoerig, P., Ludowig, E. and Knoll, I. (2008). Seeing 'where' through the ears: effects of learning-by-doing and long-term sensory deprivation on localization based on image-to-sound substitution, *Plos One* **3**(3), e1840. DOI:10.1371/journal.pone.0001840.

Proulx, M. J., Brown, D. J., Pasqualotto, A. and Meijer, P. (2014). Multisensory perceptual learning and sensory substitution, *Neurosci. Biobehav. Rev.* **41**, 16–25.

Reuschel, J., Rosler, F., Henriques, D. Y. P. and Fiehler, K. (2012). Spatial updating depends on gaze direction even after loss of vision, *J. Neurosci.* **32**, 2422–2429.

Sampaio, E., Maris, S. and Bach-y-Rita, P. (2001). Brain plasticity: 'visual' acuity of blind persons via the tongue, *Brain Res.* **908**, 204–207.

Schmidt, E. M., Bak, M. J., Hambrecht, F. T., Kufta, C. V., O'Rourke, D. K. and Vallabhanath, P. (1996). Feasibility of a visual prosthesis for the blind based on intracortical microstimulation of the visual cortex, *Brain* **119**(Pt 2), 507–522.

Shamma, S. A., Elhilali, M. and Micheyl, C. (2011). Temporal coherence and attention in auditory scene analysis, *Trends Neurosci.* **34**, 114–123.

Simon, J. Z. and Ding, N. (2010). Magnetoencephalography and auditory neural representations, *26th Southern Biomedical Engineering Conference: Sbec 2010*, **32**, 45–48.

Simpson, A. J. and Reiss, J. D. (2013). The dynamic range paradox: a central auditory model of intensity change detection, *PLoS One* **8**(2), e57497. DOI:10.1371/journal.pone.0057497.

Simpson, A. J., Reiss, J. D. and McAlpine, D. (2013). Tuning of human modulation filters is carrier-frequency dependent, *PLoS One* **8**(8), e73590. DOI:10.1371/journal.pone.0073590.

Striem-Amit, E., Guendelman, M. and Amedi, A. (2012). 'Visual' acuity of the congenitally blind using visual-to-auditory sensory substitution, *Plos One* **7**(3), e33136. DOI:10.1371/journal.pone.0033136.

Teki, S., Chait, M., Kumar, S., Shamma, S. and Griffiths, T. D. (2013). Segregation of complex acoustic scenes based on temporal coherence, *Elife* **2**, e00699. DOI:10.7554/eLife.00699.

Veraart, C., Wanet-Defalque, M.-C., Gérard, B., Vanlierde, A. and Delbeke, J. (2003). Pattern recognition with the optic nerve visual prosthesis, *Artif. Organs* **27**, 996–1004.

Wallace, M. T., Perrault, T. J., Hairston, W. D. and Stein, B. E. (2004). Visual experience is necessary for the development of multisensory integration, *J. Neurosci.* **24**, 9580–9584.

Ward, J. and Meijer, P. (2010). Visual experiences in the blind induced by an auditory sensory substitution device, *Conscious. Cogn.* **19**, 492–500.

Weiland, J. D., Liu, W. and Humayun, M. S. (2005). Retinal prosthesis, *Annu. Rev. Biomed. Eng.* **7**, 40.

Wright, B. A., Buonomano, D. V., Mahncke, H. W. and Merzenich, M. M. (1997). Learning and generalization of auditory temporal-interval discrimination in humans, *J. Neurosci.* **17**, 3956–3963.

Wright, B. A., Wilson, R. M. and Sabin, A. T. (2010). Generalization lags behind learning on an auditory perceptual task, *J. Neurosci.* **30**, 11635–11639.

Xiang, J. J., Poeppel, D. and Simon, J. Z. (2013). Physiological evidence for auditory modulation filterbanks: cortical responses to concurrent modulations, *J. Acoust. Soc. Am.* **133**, El7–El12.

Zrenner, E., Bartz-Schmidt, K. U., Benav, H., Besch, D., Bruckmann, A., Gabel, V.-P., Gekeler, F., Greppmaier, U., Harscher, A., Kibbel, S., Koch, J., Kusnyerik, A., Peters, T., Stingl, K., Sachs, H., Stett, A., Szurman, P., Wilhelm, B. and Wilke, R. (2011). Subretinal electronic chips allow blind patients to read letters and combine them to words, *Proc. Biol. Sci.* **278**(1711), 1489–1497.